

Mass Spectrometry-based Workflow for Accurate Quantification of *Escherichia coli* Enzymes: How Proteomics Can Play a Key Role in Metabolic Engineering*

Mathieu Trauchessec†§¶||, Michel Jaquinod‡§¶||, Aline Bonvalot‡§¶||, Virginie Brun‡§¶||, Christophe Bruley‡§¶||, Delphine Ropers**, Hidde de Jong**, Jérôme Garin‡§¶||, Gwenaëlle Bestel-Corre||, and Myriam Ferro‡§¶||†‡

Metabolic engineering aims to design high performance microbial strains producing compounds of interest. This requires systems-level understanding; genome-scale models have therefore been developed to predict metabolic fluxes. However, multi-omics data including genomics, transcriptomics, fluxomics, and proteomics may be required to model the metabolism of potential cell factories. Recent technological advances to quantitative proteomics have made mass spectrometry-based quantitative assays an interesting alternative to more traditional immuno-affinity based approaches. This has improved specificity and multiplexing capabilities. In this study, we developed a quantification workflow to analyze enzymes involved in central metabolism in *Escherichia coli* (*E. coli*). This workflow combined full-length isotopically labeled standards with selected reaction monitoring analysis. First, full-length ¹⁵N labeled standards were produced and calibrated to ensure accurate measurements. Liquid chromatography conditions were then optimized for reproducibility and multiplexing capabilities over a single 30-min liquid chromatography-MS analysis. This workflow was used to accurately quantify 22 enzymes involved in *E. coli* central metabolism in a wild-type reference strain and two derived strains, optimized for higher NADPH production. In combination with measurements of metabolic fluxes, proteomics data can be used to

assess different levels of regulation, in particular enzyme abundance and catalytic rate. This provides information that can be used to design specific strains used in biotechnology. In addition, accurate measurement of absolute enzyme concentrations is key to the development of predictive kinetic models in the context of metabolic engineering. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M113.032672, 954–968, 2014.

The concept of producing economically useful compounds using a “cell factory” was introduced with the advent of recombinant DNA technologies, and became fully established in the 1980s when the Food and Drug Administration approved the use of recombinant insulin (1). This concept is now largely used in the food, pharmaceutical and biotechnology industries for the production of various compounds such as polypeptides, L-threonine (2), or artemisinin, an anti-malaria drug (3) using adapted *Escherichia coli* (*E. coli*)¹ strains. As a result of increasing petrol prices and feedstock limitations, cell factories are now also used to produce chemicals from renew-

From the ‡Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA), Institut de Recherches en Technologie et Sciences pour le Vivant (IRTSV), Biologie à Grande Echelle, F-38054 Grenoble, France; §Institut National de la Santé et de la Recherche Médicale (INSERM), U1038, F-38054 Grenoble, France; ¶Université Grenoble-Alpes, F-38000 Grenoble, France; ||METabolic EXplorer, Biopole Clermont Limagne, F-63360 Saint-Beauzire, France; **Institut National de Recherche en Informatique et Automatique (INRIA), Grenoble - Rhône-Alpes, Saint Ismier CEDEX, France

Received July 19, 2013, and in revised form, January 5, 2014

Published, MCP Papers in Press, January 29, 2014, DOI 10.1074/mcp.M113.032672

Author contributions: M.T., J.G., G.B., and M.F. designed research; M.T., M.J., and A.B. performed research; M.T., M.J., V.B., C.B., D.R., H.D., G.B., and M.F. analyzed data; M.T., D.R., H.D., G.B., and M.F. wrote the paper.

¹ The abbreviations used are: DNA, Deoxyribonucleic Acid; *E. coli*, *Escherichia coli*; mRNA, messenger Ribonucleic Acid; RNA, Ribonucleic Acid; GEMs, Genome-scale models; CBM, Constraint-Based Model; ELISA, Enzyme-Linked Immuno-Sorbent assay; LC-MS, Liquid Chromatography-Mass Spectrometry; SRM, Selected Reaction Monitoring; SID, Stable Isotope Dilution; AQUA®, Accurate Quantification; QconCAT, Quantitative Concatemer; PSAQ®, Protein Standard Absolute Quantification; CCM, Central Carbon Metabolism; ATP, Adenosine Tri Phosphate; NADH, Nicotinamide Adenine Dinucleotide; NADPH, Nicotinamide Adenine Dinucleotide Phosphate; AAA, Amino Acid Analysis; NIST, National Institute of Standards and Technology; PPP, Pentose Phosphate Pathway; TCA, Tri Carboxylic acid (cycle); CV, Coefficient of Variation; LLOQ, Lowest Limit of Quantification.

The SRM data associated with this manuscript may be downloaded from Peptide Atlas using the following link: <http://www.peptideatlas.org/PASS/PASS00272>.

The MS/MS data associated with this manuscript may be downloaded from Peptide Atlas using the following link: <http://www.peptideatlas.org/PASS/PASS00273>.

able resources through fermentation, building upon advances in metabolic engineering methods (4).

Metabolic engineering consists in optimizing the genetic and regulatory processes of the cell to produce large amounts of a given chemical. Metabolic engineering requires detailed omics analyses to identify targets that can be manipulated through targeted genetic modifications to improve and/or adapt cells (5).

Various factors can affect the yield of a genetically modified cell factory, including consumption of reaction intermediates by native metabolic pathways, control of mRNA levels, regulation of small RNA or final enzyme abundances and activities. To finely tune metabolic pathways, the different levels of regulation within the cell must be identified and their contribution to overall flux control understood. Systems biology approaches aim to provide this type of information. For instance, metabolic networks can be converted to constraint-based models (CBMs), which allow metabolic fluxes to be predicted on a genome-wide scale (6) (7). As described by Feist and Palsson (8), the construction of such genome-scale models can be divided into four key steps: 1- generation of high-throughput datasets (omics data), 2- genome-wide network reconstruction, 3- conversion of network reconstruction into a CBM, and 4- application of computational models for metabolic engineering.

For quantitative data on protein levels, enzyme-linked immuno-sorbent assay (ELISA) (9) (10) has long been considered the gold standard, because of its high sensitivity and throughput; although limitations include labor intensiveness and the high cost of assay development (11). As detailed by Bantscheff *et al.* (12), proteomic approaches and instrument advances in recent years have made quantitative liquid chromatography-mass spectrometry (LC-MS) techniques a viable alternative for generating data on protein levels. Indeed, using appropriately labeled standards, MS analyses can provide highly specific and accurate data. MS is also adaptable for highly multiplexed analyses, in contrast with conventional immunoassays. Selected Reaction Monitoring (SRM), by conferring two levels of selection, filtering predefined peptide ions and specific fragment ions (13), makes it possible to detect and quantify several proteins with high sensitivity and specificity in a single assay. Combined with SRM analysis, stable isotope dilution (SID) makes absolute protein quantification feasible. SID in proteomics consists in spiking a known amount of isotopically labeled peptide or protein standards into the sample. Standards and endogenous peptide species can be distinguished in MS spectra because of their slight and predictable mass difference. In contrast, the physical and chemical properties of the different species are identical, resulting in co-elution during the LC run. Accurate information on the amount of spiked standard allows protein concentration to be deduced from the ratio of signal intensities between the heavy standard and the corresponding endogenous peptide (14).

Different standards have been described, such as labeled peptide standards (Absolute Quantification standards AQUA™), which are chemically synthesized isotope-labeled peptides. These are spiked into samples just before MS analysis (15). Alternatively, Quantitative concatemer standards (QconCAT) are artificial concatemers of labeled peptides that can be spiked into samples before trypsin digestion (16). With both these approaches, the efficiency of trypsin digestion is critical, and because prefractionation steps are often necessary, the accuracy of protein concentration measurements might be biased. Protein Standard Absolute Quantification (PSAQ™) limits these biases, because these full-length isotopically labeled standards are expected to behave in the same way as the endogenous target proteins throughout sample preparation (17). Studies by Brun *et al.* (18, 19) demonstrated new possibilities for quantitative MS, specifically in the context of clinical biomarkers detection, when targeted mass spectrometry is combined with PSAQ standards to enhance the accuracy of quantification. Although more accurate, the PSAQ strategy requires the synthesis of full-length recombinant proteins in a labeled form and may present specific difficulties.

As outlined above, to generate genome-scale models some prior biological knowledge is required. This knowledge is typically retrieved from omics experiments. We show here that an MS-based pipeline allows generating high-throughput accurate quantitative proteomics data, which can be used in combination with genome-scale models to guide choices in the design of cell factories in metabolic engineering. The study focused on *Escherichia coli* (*E. coli*), which is one of the most widely used cell factory platforms, and has been extensively studied in systems biology and used to develop very detailed mathematical models (8). We particularly targeted central carbon metabolism (CCM) because it is involved in transport and oxidation of the main carbon sources, and all the natural metabolites produced by *E. coli* are derived from 12 precursor metabolites, which are either CCM intermediates or co-factors like ATP, NADH, and NADHP (20). This makes CCM an essential crossroad for metabolic engineering (21).

We chose to use a PSAQ-like strategy to measure the concentrations of enzymes involved in *E. coli* CCM with high accuracy in a multiplexed assay. To do this, we optimized a system producing full-length ¹⁵N-labeled standards. To be fully compatible with high-throughput screening, no prefractionation steps were carried out on the *E. coli* proteome, and SRM analyses were performed using scheduled mode to allow multiplexed analyses (22). Analytical performances were tested using titration curves. The so-developed scheduled-SRM assay, monitoring 720 transitions during a single LC-SRM run, was used to accurately quantify 22 key enzymes involved in CCM in *E. coli* (Fig. 1 and Table I). Using this workflow (Fig. 2), we investigated a wild-type reference *E. coli* strain and two genetically modified strains with an increased NADPH/NADP⁺ ratio (23). Enzyme concentration measure-

TABLE I

Enzymes targeted in central carbon metabolism. Short name, full name, accession number, pathways they are involved in and molecular weight of the 22 enzymes studied are indicated. All proteins were produced in their light and heavy versions

Short name	Name	Uniprot accession number	Pathway	Molecular weight (Da)
GapA	Glyceraldehyde-3-phosphate dehydrogenase A	P0A9B2	Glycolysis	35 532
PckA	Phosphoenolpyruvate carboxykinase [ATP]	P22259	Gluconeogenesis	59 643
AceA	Isocitrate lyase	P0A9G6	Glyoxylate cycle	47 522
Zwf	Glucose-6-phosphatase 1-dehydrogenase	P0AC53	PPP	55 704
GltA	Citrate synthase	P0ABH7	TCA	48 015
MaeB	NADP-dependent malic enzyme	P76558	Gluconeogenesis	82 417
Glk	Glucokinase	P0A6V8	Glycolysis	34 723
Ppc	Phosphoenolpyruvate carboxylase	P00864	Gluconeogenesis	99 063
TktB	Transketolase 2	P33570	PPP	73 043
Pgi	Glucose-6-phosphate isomerase	P0A6T1	Glycolysis	61 530
DhsA	Succinate dehydrogenase flavoprotein subunit	P0AC41	TCA	64 422
LpdA	Dihydrolipoyl dehydrogenase	P0A9P0	Glycolysis	50 688
Mdh	Malate dehydrogenase	P61889	TCA	32 337
SucC	Succinyl-CoA ligase [ADP-forming] subunit beta	P0A836	TCA	41 393
Eda	KHG/KDPG aldolase	P0A955	Glyoxylate cycle	22 284
GpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	P62707	Glycolysis	28 556
TpiA	Triosephosphate isomerase	P0A858	Glycolysis	26 972
IcdA	Isocitrate dehydrogenase [NADP]	P08200	TCA	45 757
AckA	Acetate kinase	P0A6A3	Acetyl-CoA biosynthesis	43 290
PpsA	Phosphoenolpyruvate synthase	P23538	Gluconeogenesis	87 435
AceE	Pyruvate dehydrogenase E1 component	P06958	Glycolysis	99 668
AceB	Malate synthase A	P08997	Glyoxylate cycle	60 274

carrying an ampicillin-resistance gene, and a Profinity exact tag® (for details of the protocol, see the Bio-Rad supplier notice).

DH5α *E. coli* strains were transformed and plasmids were purified using the Nucleospin® plasmid kit. Presence of the fragment was checked by double digestion and analysis after electrophoresis on a 0.8% agarose gel. Fragments were then sequenced (GATC biotech, Konstanz, Germany). *E. coli* expression strains were stored at –80 °C in LB broth containing 20% glycerol (v/v).

Study Strains and Culture Conditions—Genetically modified and wild-type *E. coli* strains were obtained from Aurio *et al.* (23), their genotypes are described in Table II. Cultures were performed in minimal media containing 5 g/L glucose (see below). Cells were harvested during the exponential growth phase, based on OD at 600 nm.

Production and Purification of Full-length ¹⁵N-labeled Standards—**Media and Growth Conditions**—BL21-DE3 *E. coli* strains were grown on minimal medium (32 mM K₂HPO₄, 31 mM KH₂PO₄, 28.5 mM C₆H₈O₇, 4.1 mM MgSO₄, 0.3 mM CaCl₂, 14 μM ZnSO₄, 12 μM CuCl₂, 118 μM MnSO₄, 34 μM CoCl₂, 16 μM H₃BO₃, 2 μM NaMoO₄; pH 6.8). Nitrogen was only provided as (NH₄)₂SO₄ (7.5 mM final concentration). Isotopically labeled (¹⁵NH₄)₂SO₄ was used to produce labeled proteins.

For pre-culture, minimal media containing 10% LB was inoculated with the over-expressing strains of interest. Culture in minimal media was inoculated with pre-cultures to obtain an initial OD₆₀₀ of ~ 0.2. Protein expression was induced with 1 mM IPTG, and cells were grown to OD₆₀₀ ~ 4.

Cultures were centrifuged 5 min at 8000 g, supernatant was discarded, and cells were washed in the same volume of fresh PBS. Cells were resuspended in 100 mM KPO₄, pH 7.6 buffer for sonication. Lysates were centrifuged 10 min at 12,000 g and supernatant was filtered through 0.22 μm membranes.

Protein Purification Conditions and Verification of ¹⁵N isotope Incorporation—Proteins were purified using 5 ml Profinity eXact car-

tridges (Bio-Rad®) according to the supplier protocol. The duration of incubation in sodium fluoride (100 mM, pH 7.6) was adapted for each protein to optimize tag cleavage. For the proteins expressed here, between 2 h at room temperature and 12 h at 4 °C were necessary.

Purified proteins were either stored lyophilized or in 25 mM NH₄HCO₃ 10% glycerol (v/v) at –80 °C.

¹⁵N incorporation into recombinant proteins was determined on peptides after digestion of ¹⁵N labeled proteins, as described below. LC-MS/MS analysis was performed in enhanced resolution data acquisition mode on a 4000 QTRAP instrument (ABSciex, Foster City, CA), as described below. Simulated isotopic distributions with varying ¹⁵N incorporation were generated for selected ¹⁵N tryptic peptides using Isopro® software (<https://sites.google.com/site/isoproms/home>). The experimental isotope incorporation rate was determined by comparing simulated and experimental isotopic distributions using a least-squares regression method (supplemental Data File S2).

Amino Acid Analysis (AAA) for Standard Titration—Standards were calibrated using the AAA-MS (Amino Acid Analysis coupled to Mass Spectrometry) approach described by Louwagie *et al.* (26). Briefly, the protein standard (1 μg, previously titrated by UV absorbance at 280 nm) was mixed with 1 μg of NIST bovine serum albumin standard. Then microwave-assisted acidic hydrolysis was performed using the Discover protein hydrolysis apparatus (CEM, Matthews, NC).

Samples were solubilized in 20 μl of 50% acetonitrile, 0.1% formic acid. The autosampler from an HPLC-Ultimate 3000 (Dionex, Voisins Le Bretonneux, France) was used to inject 10 μl of each sample (in a 20 μl injection loop) at a flow rate of 6 μl/min (mobile phase: 50% acetonitrile, 0.1% formic acid) directly into the nanosource of an LTQ-Orbitrap XL (Thermo Fisher Scientific). MS signal intensity was averaged for each amino acid over 2.5 min. The intensity of each amino acid in the standard and the analyte were used to calculate quantities. For more details, see (26).

Trypsin and LysC Digestion—Proteins or *E. coli* lysates were solubilized in 8 M urea, 50 mM NH₄HCO₃ buffer, reduced for 30 min at

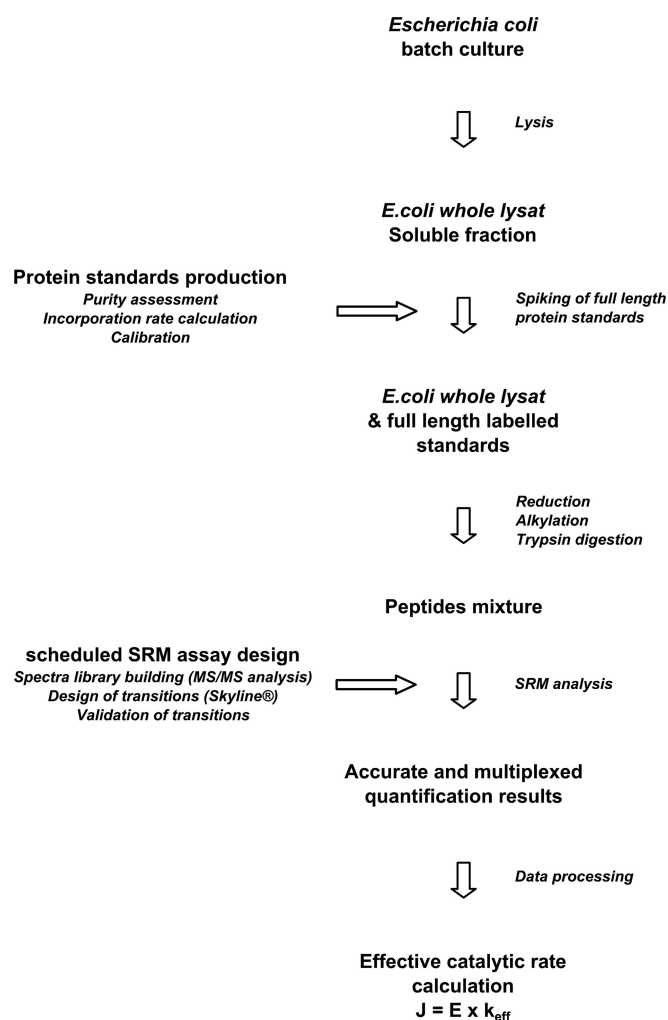


FIG. 2. Accurate quantification workflow to analyze enzymes involved in *Escherichia coli* central metabolism. Stable isotope dilution combined with an MS-based analytical strategy was developed using full-length ^{15}N labeled standards to ensure accuracy. To ensure a straightforward workflow, no prefractionation steps were carried out. Selected Reaction Monitoring (SRM) was used for optimal selectivity and sensitivity.

room temperature with 10 mM DTT and alkylated for 45 min at room temperature in the dark with 100 mM iodoacetamide. Samples were diluted with 50 mM NH_4HCO_3 to reach a final urea concentration of 2 M. LysC digestion was performed for 2 h at 37 °C at an enzyme/protein ratio of 1/100 (w/w). Samples were diluted once again with 50 mM NH_4HCO_3 to reduce urea to 0.5 M final concentration before performing trypsin digestion for 2 h at 37 °C at an enzyme/protein ratio of 1/20 (w/w). Peptide samples were finally desalted on Sep-Pack tC18 1cc vacuum cartridges, and resolubilized in a solvent containing 4% acetonitrile, 0.1% formic acid.

MS and MS/MS Experiments—MS and MS/MS analyses were performed on a hybrid triple quadrupole/ion trap mass spectrometer (4000 QTRAP; ABSciex, Les Ulis, France). Liquid chromatography (LC) separation was performed on an ultimate 3000 LC-chromatography system (Dionex, Voisins le Bretonneux, France) coupled to a Kinetex XB-C18 column, 2.1×150 mm, $2.6 \mu\text{m}$, 100Å (Phenomenex, Torrance, CA 90501). Peptides were separated using a linear 4% to 45% acetonitrile gradient over 30 min at a flow rate of 50 $\mu\text{l}/\text{min}$.

TABLE II

E. coli strains studied. Genotypes of the strains studied. Four genes were deleted in NA23 strain to induce an NADPH/NADP⁺ imbalance [23]. NA176 strain resulted from the adaptive evolution of NA23, leading to the appearance of a new pathway for NADPH reoxydation through mutations in genes *nuoF* and *rpoA*

Strain	Genotype
MG1655	Wild-type
NA23	MG1655 $\Delta pgi \Delta udhA \Delta qor \Delta edd$
NA176	MG1655 $\Delta pgi \Delta udhA \Delta qor \Delta edd \text{ } nuoF^+ \text{ } Km \text{ } rpoA^+ \text{ } Cm$

pgi: Phospho glucose isomerase.

edd: Entner Doudoroff dehydratase.

qor: Quinone oxydoreductase.

udhA: Soluble transhydrogenase.

nuoF: NADH:ubiquinone oxidoreductase subunit F.

rpoA: RNA polymerase, alpha subunit.

For MS/MS experiments, peptides were analyzed on a 4000 QTRAP hybrid triple quadrupole mass spectrometer (ABSciex, Les Ulis, France) operated using a Turbo V source in Electrospray Ionisation (ESI) mode, operated in "Information Dependent Acquisition" (IDA) mode under Analyst version 1.5.1 software (AB SCIEX). A precursor ion scan between m/z 400–1400 was performed as a survey scan for the IDA method. Enhanced product ion (EPI) spectra were acquired with a scan speed of 4000 amu/s using a dynamic fill time for optimal MS/MS and rolling collision energy settings. Precursor and product ion accuracies were 1.2 and 0.6 Da, respectively, which is in line with the specifications of the 4000 QTRAP.

Data were acquired in positive mode with the ion spray voltage at 5500 V, curtain gas 15 (arbitrary units), interface heater temperature 350 °C. Collision exit, entrance and declustering potential were set to 27, 12, and 50 volts, respectively.

For SRM assays, collision energy was calculated using a linear equation based on the manufacturer's recommendations:

$$\text{CE (volts)} = 0.44 \times m/z + 4 \text{ for doubly charged precursors}$$

$$\text{CE (volts)} = 0.5 \times m/z + 5 \text{ for triply charged precursors}$$

During SRM assay for method development, the mass spectrometer was operated in SRM mode, with a dwell time of 20 ms and a maximum of 100 transitions. After method validation, the instrument was operated using scheduled SRM mode (retention time windows: 60 s; target scan time: 2.5 s). A 30 s base width was estimated, and twelve points were acquired per chromatographic peak.

Design of SRM Assays and MS/MS Identifications—A set of twenty-two proteins was selected from among proteins involved in the central metabolic pathways in *E. coli* (Fig. 1). Each purified protein standard, in its labeled and unlabeled forms, was reduced, alkylated and digested, according to the protocol described above. LC-MS/MS analysis was performed on a 4000 QTRAP mass spectrometer. Peak lists (mgf files) were generated using Analyst version 1.5.1 software (AB SCIEX). MS/MS spectra were assigned to peptides using a sequence database search strategy. Mascot (version 2.4) was used as the search engine. We generated a home-made *E. coli* K12 protein sequence data bank (12548 sequences) from UniprotKB (<http://www.uniprot.org>, release July 28, 2011) to retrieve proteins with 83333 *Escherichia coli* taxonomy (strain K12) (<http://ebi3.uniprot.org/uniprot/?query=taxonomy%3a83333&format=>). Trypsin was set as the enzyme and 1 missed-cleavage was allowed; cystein carbamidomethyl was set as a fixed modification whereas mono and di-oxidation of methionine was set as a variable modification. Mass tolerance for precursors and fragments ions was 1.2 and 0.6 Da

respectively. Peptide charge was set at 1, 2 and 3 and instrument was ESI-TRAP. Identification results file (.dat) may be downloaded from Peptide Atlas repository (<http://www.peptideatlas.org/PASS/PASS00273>). MS/MS data were used to retrieve the best peptides for each of the 22 purified proteins in order to build SRM transition lists. Consequently, no specific cut-off score was set. Mascot results files were used to build Skyline® libraries. Lists of SRM transitions for both light and heavy versions of selected peptides were generated using Skyline® software (27). Proteotypic peptides were selected, and the sequence was confirmed to be unique using the BLAST program. In line with Jaquinod *et al.* (28), five cysteine-containing peptides were retained in the transition list. In contrast, methionine-containing peptides were not included. The retention times for the selected peptides were examined, and peptides were selected to provide a homogeneous distribution across the gradient. This resulted in selection of at least three detectable proteotypic peptides for each protein targeted in this study. For each selected peptide, three transitions were selected. The higher fragments in the acquired MS/MS spectra were preferentially chosen for setting up transitions, unless a similar transition exists for another protein. Finally, SRM transitions were validated by verifying that the retention time was identical for both heavy and light versions. Transition lists may be downloaded from Peptide Atlas repository (<http://www.peptideatlas.org/PASS/PASS00272>).

Quantitative Analysis—Quantitative SRM analyses were performed using MultiQuant software (version 1.2, AbSciex). The MultiQuant value for noise levels was 40%, and 2 min for the base-line subtraction window. All data were manually inspected to ensure correct peak detection and accurate integration. Data may be downloaded from Peptide Atlas repository (<http://www.peptideatlas.org/PASS/PASS00272>).

The heavy area/light area (H/L) ratio was determined for each SRM transition, and peptide ratios were calculated by averaging ratios for all the transitions of a given peptide. The ratios obtained for all the different proteotypic peptides from a given protein were averaged to determine the protein ratio. Standard deviation values and CVs were calculated at the peptide and protein levels considering the technical and the biological replicates (see [supplemental Data File S8](#)). Absolute quantification was obtained from the heavy/light protein ratio multiplied by the amount of standard spiked into the sample at the beginning of the pipeline. The resulting enzyme quantifications were converted to concentrations ($\text{mmol}_{\text{enzyme}} \cdot \text{ml}_{\text{cyto}}^{-1}$) and number of protein copies per cell (copies/cell) based on literature data (29) (30), as described in [supplemental Data File S3](#).

Calculation of (differences in) Effective Catalytic Rates—In steady-state growth conditions, the flux J ($\text{mmol}_{\text{metabolite}} \cdot \text{ml}_{\text{cyto}}^{-1} \cdot \text{h}^{-1}$) through a metabolic reaction can be written as $J = E \cdot k_{\text{eff}}$, where E is the enzyme concentration ($\text{mmol}_{\text{enzyme}} \cdot \text{ml}_{\text{cyto}}^{-1}$) and k_{eff} is the effective or apparent catalytic rate ($\text{mmol}_{\text{metabolite}} \cdot \text{mmol}_{\text{enzyme}}^{-1} \cdot \text{h}^{-1}$), defined similarly as in two recent papers (24) (25). k_{eff} was calculated from measurements of E and J for each enzyme in the three strains studied, as explained in [supplemental Data File S3](#). Note that fluxes in the NA23 and NA176 strains relative to the MG1655 reference strain, J/J_{MG} , can be split into the product of the relative enzyme concentrations E/E_{MG} and the relative effective catalytic rates $k_{\text{eff}}/k_{\text{eff-MG}}$: $J/J_{\text{MG}} = (E/E_{\text{MG}}) \cdot (k_{\text{eff}}/k_{\text{eff-MG}})$. Flux and enzyme concentration ratios were determined from the data for both the NA23 and NA176 strains, and relative effective catalytic rates were calculated by dividing these ratios.

RESULTS

In Cell Production of ^{15}N Labeled Protein Standards—The aim of this study was to use a proteomics quantitative pipeline, to ensure accurate data for analyzing the contributions of

enzyme concentrations and effective catalytic rates to observed flux differences between strains and for building predictive kinetic models. To begin, central metabolism pathways in *E. coli* were chosen, including glycolysis, gluconeogenesis, the glyoxylate shunt, the pentose phosphate pathway (PPP) and the tricarboxylic acid pathway (TCA) (Fig. 1). We chose to target key enzymes in each pathway, selecting glyceraldehyde-3-phosphate dehydrogenase (GapA), involved in glycolysis; glucose-6-phosphate dehydrogenase (Zwf), part of the PPP; and citrate synthase (GltA) from the TCA pathway, among others (Table I).

Twenty-two full-length ^{15}N labeled standards were produced in transformed *E. coli* and purified. To ensure that labeled standards were produced in sufficient quantities at minimal substrate cost, growth conditions were optimized to minimize $(\text{NH}_4)_2\text{SO}_4$ concentration. The purification system was chosen to ensure fast and efficient target purification. With the Bio-Scale™ Mini Profinity eXact™ cartridge, purification and tag cleavage are performed in a single step, which reduces multi-step losses (31). SDS-PAGE analysis confirmed the absence of major contaminants or degradations ([supplemental Data File S4](#)). For each ^{15}N standard, it was also necessary to check the isotope incorporation rate. This is critical for generating accurate data. Incorporation rates were evaluated using Isopro® software and found to be 98.5% for all peptides ([supplemental Data File S2](#)). This agrees well with the purity of the $(^{15}\text{NH}_4)_2\text{SO}_4$ used (98.5%). Light versions of standards were also produced to establish titration curves. Standard titration was also a critical step because it determines assay accuracy. To limit the amount of standards and time required, AAA-MS was performed, as described by Louwagie *et al.* (26). In contrast with $[^{13}\text{C}_6, ^{15}\text{N}_2]$ -L-lysine and $[^{13}\text{C}_6, ^{15}\text{N}_4]$ -L-arginine standards, full-length ^{15}N labeled standards can be titrated using all twenty amino acids; amino acids which are known to undergo chemical modifications were nonetheless excluded. For instance, asparagine and glutamine are subject to deamidation, producing aspartate or glutamate; whereas cysteine, methionine, and tryptophan may be modified through uncontrolled oxidation processes. We therefore chose five amino acids to determine the concentration of the standards produced, these were: alanine, isoleucine/leucine, phenylalanine, proline, and valine. Results are expressed as the mean concentration for these five amino acids. Each standard was assayed four times in technical replicates. The coefficients of variation (CVs) calculated for these assays were between 4 and 12% ([supplemental Data File S5](#)).

For each standard, between 2 and 5 milligrams of protein were obtained per 100 ml culture. We therefore set up a pipeline to produce large amounts of labeled protein standards while limiting substrate costs. Standards were stored at -80°C prior to further analyses.

Designing a Scheduled SRM Assay for Key Enzymes in *E. coli* Central Carbon Metabolism—To set up the SRM assay

for key enzymes involved in *E. coli* CCM, a list of peptides and fragment ions to be monitored must first be carefully defined to ensure the specificity of analyses. Skyline® software (27) was recently developed to automate this step. This software requires two inputs: protein sequences and spectral libraries for the target proteins. LC-MS/MS analyses were performed after LysC and trypsin digestion of light and heavy proteins, and MS/MS data were used to create the Skyline® library. SRM methods were automatically generated, with three proteotypic peptides per protein, except Eda, which is a small protein producing only one detectable peptide. Three transitions were monitored per peptide. All the peptides selected were proteotypic peptides, *i.e.* of unique sequence in the *E. coli* database and with good ionization properties (32). The final list for the SRM assay consisted of 720 SRM transitions (supplemental Data File S6).

To monitor these 720 SRM transitions, scheduled SRM analyses (22) were performed. Parameters were carefully tuned to optimize the limit of detection for all transitions. It was necessary to find a trade-off between the number of transitions to be monitored, peak resolution and sensitivity of the analysis. We chose to analyze samples on a micro-LC system with a flow rate at 50 μ l/min. This configuration was chosen because it results in highly reproducible retention times (\sim 10 s shift) allowing a retention time window of 60 s to be used with confidence. Target scan time was set to 2.5 s, allowing \sim 12 points per chromatographic peak, which is the peak definition recommended for quantification (33).

Thus, chromatographic conditions were optimized to ensure highly multiplexed capabilities, making it possible to monitor 720 transitions during a single 30-min effective gradient. Thus, 22 enzymes could be effectively and accurately quantified in a single analysis.

Assay Validation in Terms of Accuracy, Precision, and Linearity Range—The analytical performances of the quantification method were checked using titration curves. This involved adding increasing amount of light standard proteins to an *E. coli* lysate while adding a constant amount of heavy standard to each sample. The light version of proteins was added in amounts ranging from the lowest endogenous amount, to assess signal response linearity from basal to the highest potential levels, as advocated by the Food and Drug Administration. (<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070107.pdf>).

A preparation workflow was applied, including LysC digestion in denaturing conditions (2 M urea), followed by trypsin digestion. No prefractionation steps were applied. SRM analyses were performed to determine the heavy/light ratio and thus estimate the amount of light protein present in samples (supplemental Data File S7). Estimated light protein amounts were plotted against the quantities added. All titration points were performed as full-process triplicates. This experiment thus allows us to determine three key parameters of the assay: 1. accuracy of the method, which is indicated by the

slope of the curve; 2. precision of the method, which is indicated by the error bars corresponding to the CVs for technical replicates; 3. the range of the linear response (R^2). The results indicate that our method is accurate, precise and that the signal response is linear over the range tested (Fig. 3) (Table III). The slope of titration curves was between 0.82 (Eda) and 0.98 (PpsA), and in all cases the linear regression coefficient was higher than 0.998. The median CV for technical replicates was 2.9%, with 90% of replicates having a CV of less than 4.1%. It can be noted that, for a given protein, inter-peptide CVs were less than 20% (Supplemental data 7; note that the CV could not be determined for Eda because only one peptide was monitored).

The LLOQ is an important measure of performance as it defines the lowest analyte concentration that can be accurately measured; it thus represents a genuine measure of how sensitive an assay is. A signal-to-noise ratio of 10 is frequently used to define the LLOQ (34). However, with scheduled SRM mode, background noise is extremely low, and there is no easy and/or practical way to accurately measure co-eluting noise. We therefore chose, in line with Kuzyk *et al.* (35) (36), to determine the LLOQ of the assay empirically. Thus, the lowest analyte concentration that can be measured with a CV < 20%, and an accuracy between 80 and 120% in the linear response range is defined as the LLOQ. LLOQs were defined for each assay using data from technical triplicates for each concentration point in titration curves. The lowest point of each titration curve was the endogenous level present in the lysate (from MG1655, NA0023 or NA0176 strain), corresponding to the lowest abundance of target protein. At endogenous levels, CVs above 20% were only reached for Eda, Glk and TktB (34%, 33%, and 34%, respectively). The LLOQ considered for these three proteins was therefore the amount of light protein spiked into point 1 (P1) for titration curves (supplemental Data File S7). The LLOQs determined for each of the twenty-two protein assays are indicated in Table III.

The titration results demonstrate that our analytical pipeline is accurate, precise and linear; we therefore went on to use the labeled standards described and the associated SRM assay to measure enzyme concentrations in selected *E. coli* strains.

Absolute Quantification of Twenty-two Proteins in MG1655 *E. coli* Lysates and in Genetically Modified Strains: Assessing Workflow Efficiency—To verify the benefits of using labeled protein standards to determine *E. coli* enzyme concentrations, we first applied our SRM assay to the MG1655 wild-type reference *E. coli* strain (Fig. 2). Twenty-two standards were spiked into *E. coli* lysate. The rapid sample preparation workflow, consisting in LysC followed by trypsin digestion, was then applied, without sample decomplexification. Scheduled SRM analyses were performed as described above. Heavy/light ratios were determined, and the endogenous concentrations of each protein were calculated. Three biological

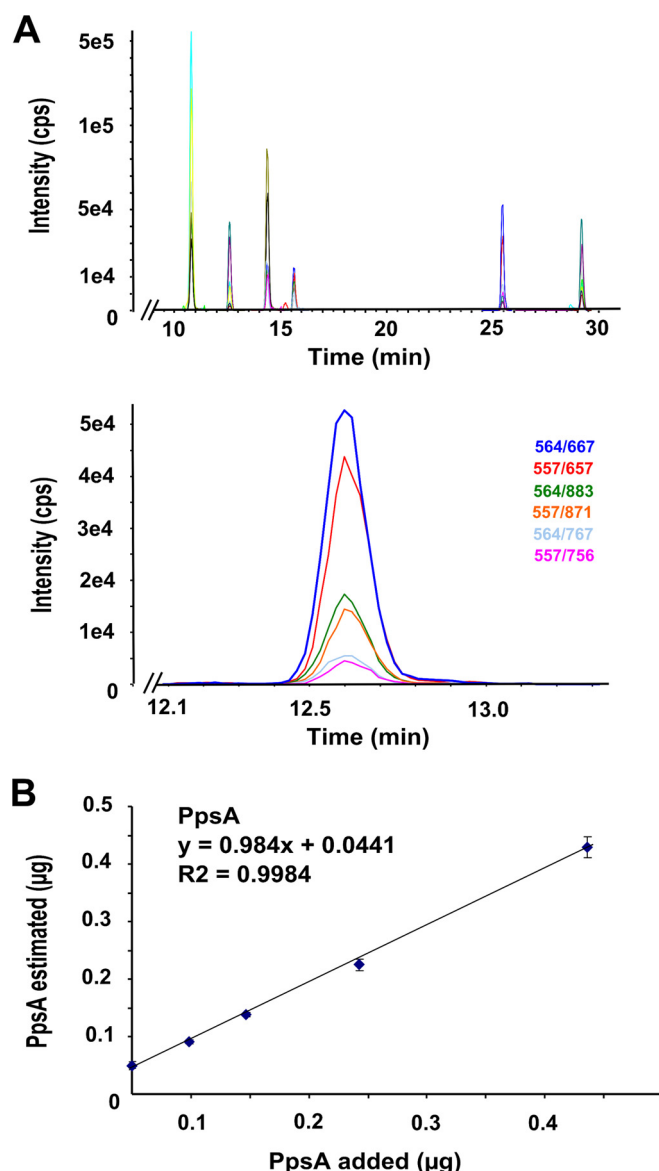


FIG. 3. Analysis of PpsA, using full-length ^{15}N labeled standard and scheduled SRM analysis. A, Total extracted ion chromatogram for PpsA. Six peptides were monitored for this enzyme, with at least three transitions per peptide. Zoom shows peptide IEDVPQEQR (retention time 12.6 min) m/z of light peptide is 557, and m/z of heavy peptide is 564. Three transitions were monitored for each version of the peptide. B, Titration curve for protein PpsA, monitoring 6 peptides and at least 3 transitions per peptide. Increasing amounts of light protein were added to samples along with a constant amount of labeled standard. The preparation workflow was applied, combining LysC and trypsin digestion, without prefractionation steps. SRM analyses were performed to determine the heavy/light ratio and thus estimate the light protein concentration. Estimated amounts of light protein were plotted against the quantities added. All titration points were performed in full-process triplicates. Linear regression was applied, and coefficients of variation are a measure of reproducibility between full-process technical triplicates.

samples were prepared for each strain, and each biological sample was analyzed twice in full-process technical replicates.

Absolute quantification was obtained for the 22 enzymes. The least abundant protein, PpsA, was quantified at 491 copies per cell, although the most abundant, GapA, was measured at 21,119 copies per cell. The biological CVs for these concentrations were 1.8% and 4.7%, respectively (Table IV), whereas technical CVs never exceeded 7%. For absolute quantification of the wild-type *E. coli* (MG1655) strain, we scanned 720 transitions during a single 60 min LC-SRM analysis (30 min effective gradient) (supplemental Data File S8). The results obtained for the wild-type strain confirmed that this workflow is robust, efficient and has a high multiplexing potential, while generating accurate, high-throughput data.

In light of these results, the pipeline was used to measure protein concentrations in two strains of biotechnological interest, both genetically engineered to produce high levels of NADPH (NA23 and NA176). NADPH is involved in many biosynthesis pathways, making it a potentially limiting cofactor when synthetic pathways are engineered. As a consequence, strains producing high levels of NADPH are interesting for metabolic engineering purposes (37). The gene *pgi*, coding for glucose-6-phosphate-isomerase, is deleted in the NA23 and NA176 strains. Hence, we determined the abundance of 21 of the 22 enzymes studied in the wild-type strain. The absolute quantities obtained from biological triplicates were expressed as a number of copies of each protein per cell (Table IV and Fig. 4). The results presented in Fig. 4 distinguish between the most abundant enzymes in each of the three strains, and make it possible to compare variations in abundance between strains. The dynamic range for these measurements was around 10^3 , because the least abundant protein, PpsA, was estimated to be present at 285 copies per cell (CV 0.7%) in strain NA176, although the most abundant protein, AceA, was estimated to be present at 58,843 copies per cell (CV 9.2%). The coefficients of variation for technical replicates never exceeded 15% for these assays and 90% of technical CVs were below 2.7% (supplemental Data File S8). For instance, for strain NA23 (biological replicate 1), the copy numbers per cell measured for Eda and AceA were 314 ± 21 and $52,650 \pm 684$, with technical CVs of 6.8% and 1.3%, respectively. This confirms the precision of the methodology right across the dynamic range.

We conclude from these results, using both technical and biological triplicates, that the workflow developed is reproducible, and can be used to generate accurate and multiplexed data at high-throughput.

Determining Effective Catalytic Rates Using Accurate Quantitative Proteomics Data and Metabolic Flux Measurements—One of the main objectives of metabolic engineering is to change metabolic flux distributions in microorganisms so as to optimize the production of compounds of biotechnological interest (38). Metabolic fluxes are controlled at different levels,

TABLE III

Titration results for the twenty-two enzymes targeted, and LLOQ estimation for each assay. Titration was performed in *E. coli* lysates for the twenty-two target proteins. At least three proteotypic peptides were used per protein, except for *Eda* for which only one proteotypic peptide was available. Three transitions were used per peptide. Each of the titration curves was fitted by linear regression; slope and linear regression coefficient were calculated. Standard deviations were calculated based on three full-process technical replicates. The LLOQ indicated were defined based on the lowest analyte concentration that could be measured with a CV < 20% (CVs for full-process technical replicates), and an accuracy between 80% and 120% in the linear response range

Protein	Accuracy ^a	Endogenous quantity ^b (μg/100 μg lysate)	Linearity ^c	Precision ^d (at mid-range)	LLOQ (μg/100 μg lysate)
<i>MG1655 lysate</i>					
LpdA	0.923	0.398	0.9994	0.6	0.414
SucC	0.919	0.196	0.9996	0.9	0.213
Mdh	0.862	0.347	0.9997	0.9	0.329
Zwf	0.940	0.061	0.9995	0.9	0.063
GltA	0.905	0.341	0.9996	1.7	0.291
PckA	0.930	0.056	0.9999	0.9	0.050
AceA	0.868	1.104	0.9967	1.7	0.913
MaeB	0.895	0.085	0.9998	1.3	0.081
GpmA	0.926	0.206	0.9993	1.3	0.212
AceB	0.901	0.403	0.9994	1.2	0.333
<i>NA0023 lysate</i>					
Pgi	0.910	0.003	0.9998	2.9	0.061
GapA	0.954	0.525	0.9998	1.2	0.528
Ppc	0.926	0.130	0.9999	0.7	0.128
AckA	0.959	0.043	1.0000	2.0	0.043
IcdA	0.908	1.111	0.9995	0.4	1.068
TpiA	0.970	0.117	1.0000	0.5	0.115
AceE	0.951	0.218	1.0000	1.3	0.220
<i>NA0176 lysate</i>					
DhsA	0.947	0.152	0.9987	3.0	0.169
Eda	0.829	0.024	0.9997	4.4	0.050
Glk	0.928	0.012	0.9997	5.4	0.036
TktB	0.968	0.044	0.9995	3.3	0.131
PpsA	0.984	0.044	0.9984	2.2	0.050

^a Accuracy: slope of titration curve.

^b Endogenous quantity: y-intercept from linear regression.

^c Linearity: correlation coefficient (R^2).

^d Precision: intertechnical replicate CVs.

including enzyme concentrations, kinetic parameters defining enzyme properties (K_m and k_{cat} in irreversible Michaelis-Menten kinetics), and the concentrations of substrates, products, and co-factors (39). We used a simple phenomenological model of metabolic fluxes, separating the effects of the abundance and the catalytic rate of the enzymes. The model makes the common assumption that metabolic fluxes are proportional to enzyme concentrations (40). Thus, $J = E \cdot k_{eff}$ where J denotes metabolic flux, E enzyme concentrations, and k_{eff} the so-called effective catalytic rate. The effective catalytic rate describes how the maximum catalytic rate k_{cat} is modulated by concentrations of substrates and products as well as co-factors regulating the catalytic action of the enzyme. As a consequence, k_{eff} is not a constant enzyme property, like k_{cat} , but depends on both enzyme properties and condition-varying metabolite concentrations. By definition, the effective catalytic rate is smaller than the maximum catalytic rate, i.e., $k_{eff} \leq k_{cat}$ (supplemental Data File S3). The model $J = E \cdot k_{eff}$ can be used even when exact, *in vivo* values for the kinetic parameters are missing (which is often the case).

Absolute values were available for two of the three variables in the flux equation for the network of Fig. 1, for all three

strains: the metabolic flux distribution, as described by Auriol *et al.* (23), and the enzyme concentrations from the SID-scheduled SRM (scSRM) analysis presented in this study. The effective catalytic rate could thus be computed from $k_{eff} = J/E$, after converting the experimental data to comparable units (supplemental Data File S3). This idea can be used to gain a deeper insight into how flux differences across strains are related to relative concentrations and effective catalytic rates of enzymes, by comparing the wild-type reference strain MG1655 with strains NA23 and NA176. The metabolic flux in NA23 or NA176 relative to MG1655 (J/J_{MG}) equals the product of the relative enzyme concentrations (E/E_{MG}) and the relative catalytic rates (k_{eff}/k_{eff-MG}). For example, a twofold increase in metabolic flux in strain NA23 ($J_{NA23}/J_{MG} = 2$) could be because of a fourfold increase in enzyme concentration ($E_{NA23}/E_{MG} = 4$) and a twofold decrease in its effective catalytic rate ($k_{eff-NA23}/k_{eff-MG} = 0.5$).

We measured the relative metabolic fluxes and enzyme concentrations, and calculated effective catalytic rates for 19 out of the 21 enzymes in all three strains growing on minimal medium with glucose (PpsA and Glk were not considered, as there was no flux data available for the corresponding reactions). The results show that, for some reactions, the meta-

TABLE IV

Accurate and multiplexed quantification results for twenty-two proteins in *E. coli* lysates. The proteomics workflow was used to accurately quantify 22 key enzymes in *E. coli* central carbon metabolism during exponential growth in minimal medium supplemented with glucose. Accurate quantities determined with full-length ^{15}N labelled standards detected by scheduled SRM analyses are shown. For each strain, biological triplicates were performed, and full-process technical triplicates were run for each sample. Accurate concentrations are expressed in number of protein copies per cell. The coefficients of variation are calculated based on data for biological triplicates

		Average quantity from biological replicates					
		Number of copies per cell					
		MG1655		NA23		NA176	
		Copies per cell	CV (%)	Copies per cell	CV (%)	Copies per cell	CV (%)
Glycolyse	Glk	289*	1.8	463	16.7	119*	1.5
	PpsA	491	1.8	690	4.4	285	0.7
	Pgi	2431	2.1				
	AceE	3028	1.0	1022	10.7	1880	3.8
	TpiA	3665	2.3	2046	33.5	1673*	2.4
	GpmA	4945	1.6	5446	2.8	5325	3.8
	LpdA	5512	3.1	3804	3.8	11,301	3.5
	GapA	21,119	4.7	7605	33.3	10,749	5.9
Pentose phosphate pathway	Zwf	778	2.9	579	13.7	1091	2.7
	TktB	1019	4.3	887	26.4	268*	4.1
Entner Doudonoff pathway	Eda	2049	3.0	451*	43.9	484*	2.1
Glyoxylate shunt	AceB	3846	2.7	7140	23.4	6433	4.0
	GltA	4325	4.3	15,160	8.8	14,986	3.3
	Mdh	7137	3.8	10,943	7.8	14,683	2.7
	AceA	12,251	5.7	58,843	9.2	39,767	2.7
Krebs cycle	DhsA	2002	7.1	8290	12.3	1472	4.7
	SucC	3686	5.9	7336	2.9	4092	4.1
	IcdA	12,145	3.4	11,446	16.1	30,637	2.6
Anaplerotic pathways	PckA	525	3.3	3185	12.5	1133	4.7
	MaeB	708	4.8	1608	10.4	919	3.2
	Ppc	2668	3.6	748	49.5	2990	2.2
Acetate pathway	AckA	686	14.0	420	5.6	431*	3.2

* Absolute quantities indicated are close to but below the LLOQ determined based on titration curves.

bolic flux differences are mainly because of variations in enzyme concentrations, whereas for others they are the result of changes in the catalytic activity of the enzymes (Table V and supplemental Data File S9).

For instance, Auriol *et al.* (23) compared the NA23 strain with the wild-type strain and found a 53-fold decrease in flux through the reaction catalyzed by SucC, which generates succinate from succinyl-CoA as part of the TCA cycle (Fig. 1). We show that this change in flux results from a 3.3-fold increase in enzyme concentration and a 180-fold decrease in effective catalytic rate (Table V). This confirms that the difference in catalytic rate causes the change in metabolic flux. It would not have been possible to reach this conclusion based on flux data alone. Similar observations can be made for the reactions catalyzed by DhsA, GltA, and GpmA, for which the fluxes decreased 14, 14, 16-fold, respectively, in strain NA23 relative to the wild-type MG1655 strain. Although the enzyme concentrations increased 6.9, 5.9, 1.8-fold for these reactions, the effective catalytic rates decreased 100, 83, and 29-fold, respectively (Table V). For AceA, Auriol *et al.* observed a fivefold increase in flux. Our results indicate that this is the result of an 8.1-fold increase in enzyme concentration, and a 1.6-fold decrease in effective catalytic rate. The results observed for AceA in this study match those reported by Hua *et al.* (41), who demonstrated that deleting *pgi* led to in-

creased activation of the glyoxylate shunt. Consistent with this, IcdA abundance increased 1.6-fold, whereas its catalytic rate decreased 81-fold. This leads to a decreased flux through the TCA cycle, with substrates redirected through the glyoxylate shunt (see below).

In addition to quantifying the relative contributions of enzyme concentrations and activities to observed flux changes, the computation of effective catalytic rates may help in pinpointing flux bottlenecks in (reengineered) metabolic networks. If the effective catalytic rate is close to k_{cat} , the enzyme is operating not far from its maximum capacity. As a consequence, a further increase of flux through the reaction will require an increase of the enzyme concentration. In the NA176 strain, for instance, the flux through the glyoxylate shunt is much increased in comparison with the wild-type MG1655 strain and the effective catalytic rate of AceB is 34.7 s^{-1} . When comparing this number with the k_{cat} constant, using the compendium of Bar-Even *et al.* (42), the values of k_{eff} and k_{cat} are found to be close ($k_{cat} = 48.1 \text{ s}^{-1}$). This indicates a limiting capacity constraint in the network that could be resolved by overexpressing the enzyme. Notice that this diagnostic critically depends on the possibility to compare k_{eff} with k_{cat} , and thus on the capability to measure absolute enzyme concentrations with high precision and accuracy, like in this study.

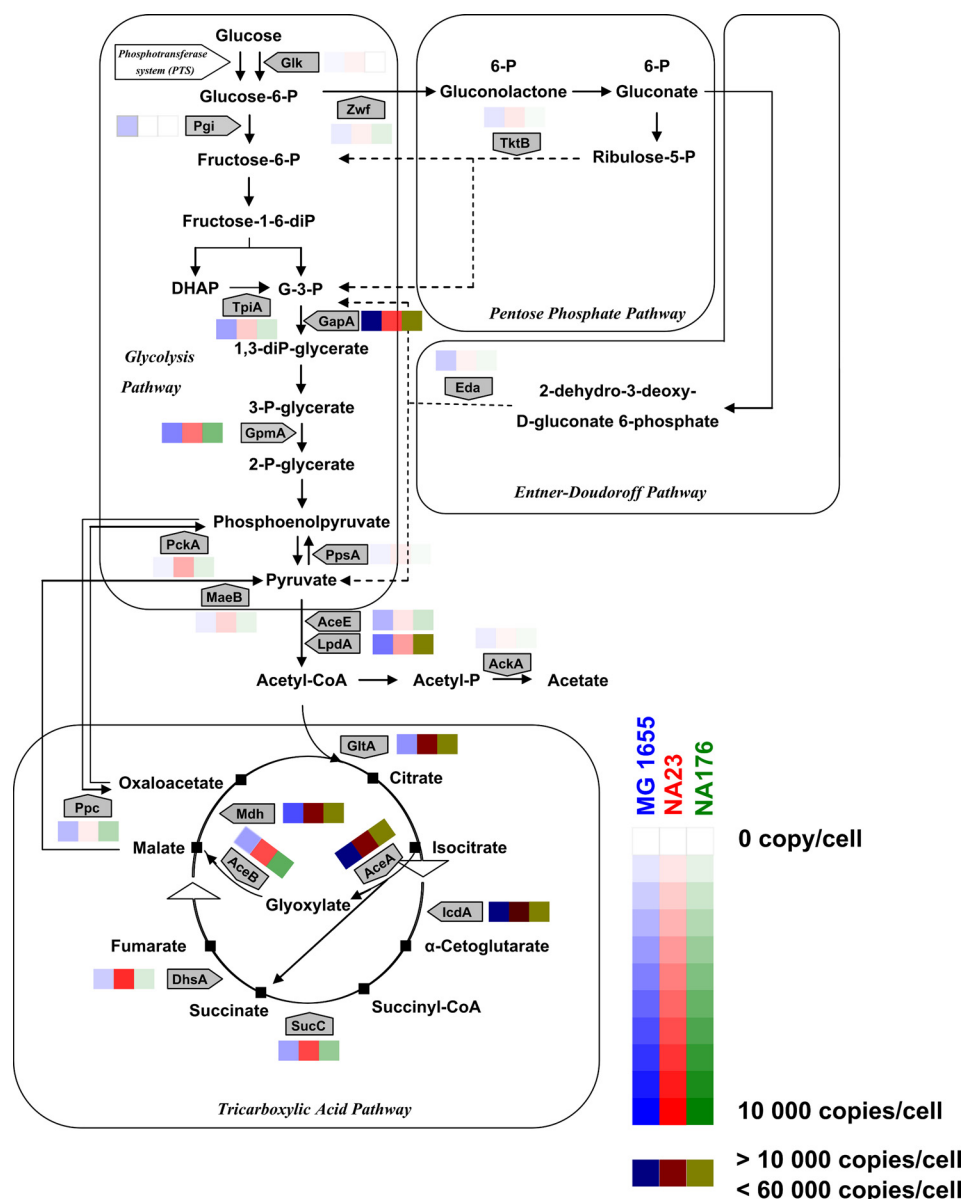


FIG. 4. *Escherichia coli* central metabolism and absolute amounts of 22 enzymes involved in these pathways. The proteomics workflow was used to accurately quantify 22 key enzymes in *E. coli* central carbon metabolism during exponential growth in minimal medium supplemented with glucose. Absolute amounts are expressed as a number of protein copies per cell for 22 proteins from the wild-type strain (MG1655), and for 21 proteins from the genetically modified strains (NA23 and NA176) from which the *pgi* gene is deleted. Results for MG1655, NA23 and NA176 strains are indicated in blue, red and green, respectively, with a color intensity code to broadly indicate number of copies per cell. This representation allows ready visualization of the most and least abundant proteins in central metabolism, and of variations between wild-type and modified strains.

Using the workflow and the labeled protein standards developed in this work, absolute enzyme concentrations can be obtained for any *E. coli* strain cultivated in any experimental conditions, and effective catalytic rates can be calculated from flux data, if available. As demonstrated by the comparison of the wild-type and mutant strains, information on the effective catalytic rate of an enzyme can be used to distinguish between two levels of regulation of metabolic fluxes: enzyme concentrations and the rates at which enzymes cat-

alyze reactions. The insights thus obtained can be used to target metabolic engineering efforts, for instance to decide whether it is worth increasing the expression of a gene coding for a specific enzyme.

DISCUSSION

This article shows the use of an absolute and multiplexed SID-MS based quantification workflow for modeling purposes, dedicated to the in-depth study of *E. coli* central car-

TABLE V

Variations in fluxes, enzyme concentrations, and effective catalytic rates between strains MG1655, NA23 and NA176. Metabolic fluxes were determined by Auriol *et al.* [23], enzyme concentrations were determined in this study. Metabolic fluxes and enzyme concentrations can be used to calculate effective catalytic rates (supplemental Data File S3 and supplemental Data File S9). The table shows the relative difference in metabolic flux, enzyme concentrations, and effective catalytic rate for 19 enzymes between the reference strain, MG1655, and strains NA23 and NA176. *pgi*, which codes for Pgi, is deleted from both NA23 and NA176 strains, while for the reactions catalyzed by PpsA and Glk no flux data were available

Strain NA23 versus MG1655			
	J_{NA23}/J_{MG}	E_{NA23}/E_{MG}	$k_{eff-NA23}/k_{eff-MG}$
AceE	0.085	0.566	0.150
TpiA	0.072	0.935	0.077
GpmA	0.064	1.846	0.035
LpdA	0.085	1.157	0.073
GapA	0.046	0.604	0.076
Zwf	5.428	1.247	4.353
TktB	1.717	1.458	1.177
Eda	0.000	0.369	0.000
AceB	4.985	3.112	1.602
GlxA	0.069	5.875	0.012
Mdh	0.110	2.570	0.043
AceA	4.985	8.050	0.619
DhsA	0.071	6.940	0.010
SucC	0.019	3.336	0.006
IcdA	0.020	1.580	0.012
PckA	0.392	10.161	0.039
Ppc	0.004	0.470	0.009
MaeB	1.329	3.808	0.349
AckA	0.000	1.025	0.000

Strain NA176 versus MG1655			
	J_{NA176}/J_{MG}	E_{NA176}/E_{MG}	$k_{eff-NA176}/k_{eff-MG}$
AceE	0.337	0.698	0.483
TpiA	0.637	0.513	1.242
GpmA	0.531	1.210	0.439
LpdA	0.337	2.304	0.146
GapA	0.393	0.572	0.688
Zwf	52.000	1.574	33.028
TktB	14.560	0.295	49.301
Eda	0.000	0.265	0.000
AceB	8.320	1.880	4.425
GlxA	0.280	3.894	0.072
Mdh	0.251	2.312	0.108
AceA	8.320	3.648	2.281
DhsA	0.250	0.826	0.302
SucC	0.157	1.248	0.126
IcdA	0.198	2.835	0.070
PckA	0.072	2.425	0.030
Ppc	0.539	1.259	0.428
MaeB	8.320	1.459	5.702
AckA	0.096	0.706	0.136

bon metabolism. The method is based on stable isotope dilution combined with MS-based analysis using Selected Reaction Monitoring (SRM). Previous studies have demonstrated that several proteins could be detected and accurately quantified with high sensitivity and specificity using peptides (43) or concatemers standards (44) coupled to SRM analysis. More especially, Carroll *et al.* (44) described the absolute quantification of 27 enzymes (and isoenzymes) of the yeast

glycolysis pathway using one 88 kDa size concatemers. In the present work, we chose an absolute quantification methodology that allows the best possible accuracy for CCM enzyme concentrations measurements. Indeed, in order to obtain highly accurate measurements, choices had to be made, concerning the type of standards, the MS analysis mode, and quality control of the overall workflow. In the context of large scale studies, AQUA and QconCAT standards are commonly used for multiplexed absolute quantification experiments. For example, a QconCAT strategy has been developed with the objective to provide absolute quantification for at least 4000 proteins of yeast (45). However, as described by Brownridge *et al.*, "Quantification is impaired if either the QconCAT or the analyte proteins are incompletely digested, such that the yield of either peptide is incomplete - indeed, this is not a problem unique to our workflow, but any quantitative approach using proteolytic digestion to generate peptides as analytes." Thus, to ensure accurate measurement of protein concentrations, stable isotope labeled proteins standards were chosen as reference. Added at the very beginning of sample preparation, those standards allow eliminating all biases and ensuring accurate quantification. We decided to use full-length isotope labeled standards to measure enzyme concentrations as accurately as possible. Considering our study model (*E. coli*), it appeared relevant to produce full-length isotopically labeled standards in *E. coli* cells, to allow high production rates. The first step was to develop a minimal growth medium limiting substrate consumption and thus production costs. Then, a purification system had to be found to allow easy and efficient production of standards. We selected a system where purification and tag cleavage were performed during a single step (31). We then verified the isotope incorporation rate by comparing precursor ion mass spectra with a simulated isotopic distribution. Finally, standards were calibrated, using an AAA-MS approach which has the advantages of requiring tiny amounts of starting material and a shorter time frame, and which has similar accuracy to standard AAA (26). This workflow can be used for straightforward and efficient production of large amounts of multiple full-length isotopically labeled standards, and is particularly suitable for soluble proteins. This system is not limited to *E. coli* proteins, and can be extended for many applications where post-translational modifications are not needed. For instance, Wang *et al.* (31) developed a similar ^{15}N labeled standard production system to quantify human ApolipoproteinE, whereas another study (46) produced ^{15}N -labeled analogues to accurately quantify DNA glycosylase.

A recent study has demonstrated that an SRM approach can quantify proteins over a dynamic range between 50 and 10^6 copies per cell (43). According to Ishihama *et al.* (47), the protein dynamic range in *E. coli* strains is ~ 4 orders of magnitude. The selectivity, sensitivity, and dynamic range capabilities conferred by SRM make this approach the most appropriate for our study. One of the main challenges was to

develop an analytical workflow allowing high-throughput, precise and accurate quantification across the whole range of protein abundances for *E. coli*. We used a micro-LC configuration to generate highly resolved chromatographic peaks, with highly reproducible retention times. This allowed us to accurately quantify 22 proteins, monitoring 720 transitions during a single 30-min analysis. With only three peptides per protein and three transitions per peptide (for both heavy and light forms), the scSRM method described here could easily be used for accurate quantification of up to 50 proteins simultaneously. The analytical performance of our workflow was assessed using titration curves, and results showed a high precision (low inter-technical and inter-peptide CVs). The precision obtained led us to conclude that with full-length standards, it should be possible to monitor only one peptide without losing accuracy. With one peptide per protein and three transitions per peptide (for both heavy and light forms), it might be possible to accurately quantify 150 proteins in a single run by scanning 900 transitions. Keseler *et al.* (48) described small molecule metabolism in *E. coli* to be dependent on 865 enzymes. This means that our technique could potentially accurately quantify 20% of *E. coli* enzymes in a single 30-min analysis.

We applied the quantification workflow to strains with increased production of NADPH, an essential intermediate involved in many biosynthesis pathways. We focused on how this imbalance affected the abundance of CCM enzymes (21). To do this, we studied the modified microorganisms developed by Auriol *et al.* (23). In strain NA23 (MG1655 $\Delta pgi \Delta edd \Delta udhA \Delta qor$) NADPH overproduction was the result of deletion of the *pgi* and *edd* genes. Auriol *et al.* showed that, to compensate for the stress because of the NADPH/NADP⁺ imbalance, two adaptive mutations in two different genes occurred to produce strain NA176 (23). The first of these, *nuoF*^{*}, occurred in the gene encoding a subunit of the water-soluble NADH:ubiquinone, leading to novel NADPH reoxidizing pathways. The second mutation, *rpoA*^{*}, occurred in the gene coding for the alpha subunit of RNA polymerase, which causes a global transcriptional rearrangement, involving the repression of genes contributing to the TCA pathway and the glyoxylate shunt (23). Results presented by Auriol *et al.* showed that, in the case of NA23 and NA176, fluxes were rerouted through the PPP, whereas the lower glycolysis pathway was activated to produce pyruvate. TCA enzymes were regulated through hyperactivation of the glyoxylate shunt, thus limiting NADPH production by isocitrate dehydrogenase (Fig. 1). This resulted in an accumulation of oxaloacetate, which explains the need to activate anaplerotic pathways to maintain the balance of TCA intermediates (21).

How can the accurate quantitative proteomics data obtained with the approach presented in this study be exploited for the analysis of the network? The absolute numbers in Table IV are interesting in themselves, as they show to which CCM enzymes the cell directs its protein synthesis invest-

ments. However, the data are even more valuable when cross-referenced with other types of omics resources.

We have combined the proteomics data with flux distributions available from experiments carried out under exactly the same conditions as in this study. This makes it possible to analyze the relative contributions of the concentration and the catalytic rate of enzymes to observed flux variations between strains. For NA23, the increased flux through the AceA reaction in the glyoxylate shunt was found to be mainly because of a higher enzyme concentration with a relatively constant effective catalytic rate. A similar result was observed for AceB. In contrast, flux reduction through *lcdA* appeared to be mainly because of a diminished catalytic rate of the enzymes. These results concur with studies describing regulation of the glyoxylate bypass operon, *aceBAK*. This operon codes for the glyoxylate shunt enzymes AceA and AceB, as well as isocitrate dehydrogenase kinase/phosphatase (AceK). AceK phosphorylates *lcdA*, to reduce its activity (49) (50). Thus, our results, showing that AceA and AceB were present in increased amounts whereas the effective catalytic rate of *lcdA* decreased, are consistent with the known co-expression of *aceA*, *aceB*, and *aceK* as well as the post-translational regulation of *lcdA* by AceK. In the NA176 strain, which results from adaptive evolution of NA23, the *lcdA* flux was higher than in NA23. This appears to be because of an increased effective catalytic rate, as the enzyme concentration remains similar in both strains. The above results validate the consistency of our strategy for calculating *in vivo* effective catalytic rates in prokaryote models from fluxes and enzymes concentrations. Recently, effective catalytic rates were also used for understanding the complex relation between the protein synthesis rate and the growth rate of the cell (24)(25).

Quantitative proteomics, through the computation of effective catalytic rates of enzymes, thus makes it possible to dissect network functioning in a novel way. In addition, this information may also be of great value in the context of metabolic engineering, as it indicates which enzymes operate close to their maximum capacity and may thus be bottlenecks in the network. For example, in the NA176 strain, where the flux is rerouted through the glyoxylate shunt, the k_{eff} value of AceB was found to be close to its maximal value, given by k_{cat} , thus suggesting a possible target for protein overexpression. Although interesting as a heuristic guideline, one should be beware of the limits of this strategy. Assays for determining k_{cat} values are performed by means of purified enzymes *in vitro*, and may therefore lead to a distorted view of the actual *in vivo* situation (51). Even more importantly, control of metabolic fluxes is a global network property and the relaxation of local capacity constraints for individual reactions is not guaranteed to increase the flux through those reactions (40). Therefore, it will be interesting to integrate accurate quantitative proteomics data into predictive models, both constraint-based models (CBMs) of the flux distribution in metabolic networks and mechanistic models including details of enzyme

kinetics. For instance, enriching CBMs with quantitative proteomics and metabolomics data, as part of IOMA (52), would increase the accuracy of the metabolic fluxes predicted. Ultimately, quantitative information on protein abundance may be most useful when combined with both flux and metabolite data in the construction of predictive kinetic models, which explicitly take into account the regulatory mechanisms of the cell.

A recent example of the latter strategy is the study by Smallbone *et al.* (53), who combined proteomics and metabolomics data on yeast glycolysis with enzyme kinetic assays to obtain *in vivo* values for k_{cat} and K_m which, when integrated into a dynamic kinetic model, allowed to refine the picture of the control of yeast glycolysis, distributed more widely over the whole system than originally thought. Another example is the data set of Ishii *et al.*, who measured in parallel enzyme and metabolite concentrations as well as fluxes and mRNA levels in CCM of *E. coli* (54). These data have been used to estimate the parameter values in a simplified model of the metabolic network, revealing among other things a number of critical issues for model identification in genome-scale kinetic models, such as theoretical and practical limits on the identifiability of parameter values (55).

A limitation of the present work is related to the fact that isoenzymes were not individually considered when analyzing effective catalytic rates. As a consequence, the effective rates of some enzymes might possibly be incorrectly calculated as fluxes might result from the action of several enzymes. However, in prokaryote cells, like *E. coli*, isoenzymes are rare because a given activity is more commonly controlled by a single protein. Among the 22 chosen enzymes, only the transketolase activity involved two isoenzymes, TktA and TktB, of which only TktB was measured. In order to obtain a reliable estimate of the overall *in vivo* contribution of enzyme abundance and effective catalytic rate to a change in flux, the abundances of both isoenzymes would have to be quantified. In order to separate the activity of each isoenzyme, *in vitro* assays can be performed as described by Smallbone *et al.* (53).

In conclusion, in this work we describe the development of a proteomics workflow allowing multiplex absolute quantification of enzymes involved in *E. coli* CCM. We analyzed the data thus obtained in an innovative way, by calculating effective catalytic rates of the enzymes for wild-type and mutant strains optimized for higher NADPH production. This provides clues about the relative contributions of enzyme concentrations and their catalytic rates to flux changes observed across different strains as well as about potential flux bottlenecks in the network. Although important in their own right, the results thus obtained also lead the way toward kinetic models of metabolism that can be used as predictive tools in systems biology and many applications in biotechnology.

Acknowledgements—We thank Dr Christine Carapito and Dr Rainer Figge for scientific discussions. Thanks to EDyP team, to Sylvie Kieffer-Jaquinod for assistance with AAA-MS analyses, to Mathilde Louwagie for assistance with AAA sample preparation, to the

METabolic EXplorer team for training and technical assistance when producing the protein standards. We are also grateful to Dr Maighread Gallagher-Gambarelli for editorial assistance.

* This project was supported by the «Association Nationale de la Recherche et de la Technologie», METabolic EXplorer (Saint-Beauzire, France), the Investissement d'Avenir Bioinformatics program (RESET project, ANR-11-BINF-0005) and the Investissement d'Avenir Infrastructures Nationales en Biologie et Santé program (ProFI project, ANR-10-INBS-08).

§ This article contains [supplemental Data Files S1 to S10](#).

‡‡ To whom correspondence should be addressed: CEA, 17 Rue des Martyrs, Grenoble 38054, France. Tel.: +334-38782549; E-mail: myriam.ferro@cea.fr.

REFERENCES

- Johnson, I. S. (1983) Human insulin from recombinant DNA technology. *Science* **219**, 632–637
- Lee, J. H., Sung, B. H., Kim, M. S., Blattner, F. R., Yoon, B. H., Kim, J. H., and Kim, S. C. (2009) Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microb. Cell Fact.* **8**, 2
- Ro, D., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943
- Lee, S. Y., Mattanovich, D., and Villaverde, A. (2012) Systems metabolic engineering, industrial biotechnology and microbial cell factories. *Microb. Cell Fact.* **11**, 156
- Nielsen, J., and Keasling, J. D. (2011) Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol.* **29**, 693–695
- Kitano, H. (2002) Systems biology: a brief overview. *Science* **295**, 1662–1664
- Sauer, U., Heinemann, M., and Zamboni, N. (2007) Genetics. Getting closer to the whole picture. *Science* **316**, 550–551
- Feist, A. M., and Palsson, B. Ø. (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* **26**, 659–667
- Voller, A., Bartlett, A., and Bidwell, D. E. (1978) Enzyme immunoassays with special reference to ELISA techniques. *J. Clin. Pathol.* **31**, 507–520
- Voller, A. (1978) The enzyme-linked immunosorbent assay (ELISA) (theory, technique and applications). *Ric. Clin. Lab.* **8**, 289–298
- Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N., and Wang, P. (2008) The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl.* **2**, 1386–1402
- Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404**, 939–965
- Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222
- Barr, J. R., Maggio, V. L., Patterson, D. G. J., Cooper, G. R., Henderson, L. O., Turner, W. E., Smith, S. J., Hannon, W. H., Needham, L. L., and Sampson, E. J. (1996) Isotope dilution–mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.* **42**, 1676–1682
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6940–6945
- Beynon, R. J., Doherty, M. K., Pratt, J. M., and Gaskell, S. J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2**, 587–589
- Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell. Proteomics* **6**, 2139–2149
- Brun, V., Masselon, C., Garin, J., and Dupuis, A. (2009) Isotope dilution strategies for absolute quantitative proteomics. *J. Proteomics* **72**, 740–749

19. Huillet, C., Adrait, A., Lebert, D., Picard, G., Trauchessec, M., Louwagie, M., Dupuis, A., Hittinger, L., Ghaleh, B., Le Corvoisier, P., Jaquinod, M., Garin, J., Bruley, C., and Brun, V. (2012) Accurate quantification of cardiovascular biomarkers in serum using Protein Standard Absolute Quantification (PSAQ™) and selected reaction monitoring. *Mol. Cell. Proteomics* **11**, M111.008235
20. Nielsen, J. (2003) It is all about metabolic fluxes. *J. Bacteriol.* **185**, 7031–7035
21. Papagianni, M. (2012) Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microb. Cell Fact.* **11**, 50
22. Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., and Domon, B. (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics* **6**, 1809–1817
23. Aurioi, C., Bestel-Corre, G., Claude, J., Soucaille, P., and Meynial-Salles, I. (2011) Stress-induced evolution of *Escherichia coli* points to original concepts in respiratory cofactor selectivity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1278–1283
24. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., and Palsson, B. Ø. (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9**, 693
25. Valgepea, K., Adamberg, K., Seiman, A., and Vilu, R. (2013) *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* **9**, 2344–2358
26. Louwagie, M., Kieffer-Jaquinod, S., Dupierri, V., Couté, Y., Bruley, C., Garin, J., Dupuis, A., Jaquinod, M., and Brun, V. (2012) Introducing AAA-MS, a rapid and sensitive method for amino acid analysis using isotope dilution and high-resolution mass spectrometry. *J. Proteome Res.* **11**, 3929–3936
27. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968
28. Jaquinod, M., Trauchessec, M., Huillet, C., Louwagie, M., Lebert, D., Picard, G., Adrait, A., Dupuis, A., Garin, J., Brun, V., and Bruley, C. (2012) Mass spectrometry-based absolute protein quantification: PSAQ™ strategy makes use of “noncanonical” proteotypic peptides. *Proteomics* **12**, 1217–1221
29. Volkmer, B., and Heinemann, M. (2011) Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One* **6**, e23126
30. Hiller, J., Franco-Lara, E., and Weuster-Botz, D. (2007) Metabolic profiling of *Escherichia coli* cultivations: evaluation of extraction and metabolite analysis procedures. *Biotechnol. Lett.* **29**, 1169–1178
31. Huang, L., Mao, X., Abdulaev, N. G., Ngo, T., Liu, W., and Ridge, K. D. (2012) One-step purification of a functional, constitutively activated form of visual arrestin. *Protein Expr. Purif.* **82**, 55–60
32. Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., and Aebersold, R. (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **25**, 125–131
33. Holcapek, M., Jirasko, R., and Lisa, M. (2012) Recent developments in liquid chromatography-mass spectrometry and related techniques. *J. Chromatogr. A* **1259**, 3–15
34. Keshishian, H., Addona, T., Burgess, M., Kuhn, E., and Carr, S. A. (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **6**, 2212–2229
35. Green, J. (1996) A practical guide to analytical method validation. *Anal. Chem.* **68**, 305A–309A
36. Kuzyk, M. A., Smith, D., Yang, J., Cross, T. J., Jackson, A. M., Hardie, D. B., Anderson, N. L., and Borchers, C. H. (2009) Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics* **8**, 1860–1877
37. Kabir, M. M., and Shimizu, K. (2003) Fermentation characteristics and protein expression patterns in a recombinant *Escherichia coli* mutant lacking phosphoglucose isomerase for poly(3-hydroxybutyrate) production. *Appl. Microbiol. Biotechnol.* **62**, 244–255
38. Stephanopoulos, G. N., Aristidou, A. A., and Nielsen, J. (1998) *Metabolic Engineering: Principles and Methodologies*. San Diego: Academic Press (Ed.)
39. Gerosa, L., and Sauer, U. (2011) Regulation and control of metabolic fluxes in microbes. *Curr. Opin. Biotechnol.* **22**, 566–575
40. Heinrich, R., and Schuster, S. (1996) *The regulation of cellular systems*. London: Chapman & Hall
41. Hua, Q., Yang, C., Baba, T., Mori, H., and Shimizu, K. (2003) Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J. Bacteriol.* **185**, 7053–7067
42. Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410
43. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R. (2009) Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806
44. Carroll, K. M., Simpson, D. M., Evers, C. E., Knight, C. G., Brownridge, P., Dunn, W. B., Winder, C. L., Lanthaler, K., Pir, P., Malys, N., Kell, D. B., Oliver, S. G., Gaskell, S. J., and Beynon, R. J. (2011) Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. *Mol. Cell. Proteomics* **10**, M111.007633
45. Brownridge, P., Holman, S. W., Gaskell, S. J., Grant, C. M., Harman, V. M., Hubbard, S. J., Lanthaler, K., Lawless, C., O’Cualain, R., Sims, P., Watkins, R., and Beynon, R. J. (2011) Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics* **11**, 2957–2970
46. Dizzardoglu, M., Reddy, P. T., and Jaruga, P. (2011) Identification and quantification of DNA repair proteins by liquid chromatography/isotope-dilution tandem-mass spectrometry using their fully ¹⁵N-labeled analogues as internal standards. *J. Proteome Res.* **10**, 3802–3813
47. Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**, 102
48. Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., Peralta-Gil, M., and Karp, P. D. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33**, D334–7
49. Yamamoto, K., and Ishihama, A. (2003) Two different modes of transcription repression of the *Escherichia coli* acetate operon by IclR. *Mol. Microbiol.* **47**, 183–194
50. Rittinger, K., Negre, D., Divita, G., Scarabel, M., Bonod-Bidaud, C., Goody, R. S., Cozzzone, A. J., and Cortay, J. C. (1996) *Escherichia coli* isocitrate dehydrogenase kinase/phosphatase. Overproduction and kinetics of interaction with its substrates by using intrinsic fluorescence and fluorescent nucleotide analogues. *Eur. J. Biochem.* **237**, 247–254
51. Garcia-Contreras, R., Vos, P., Westerhoff, H. V., and Booger, F. C. (2012) Why in vivo may not equal in vitro - new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *FEBS J.* **279**, 4145–4159
52. Yizhak, K., Benyamini, T., Liebermeister, W., Rupp, E., and Shlomi, T. (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **26**, i255–60
53. Smallbone, K., Messiha, H. L., Carroll, K. M., Winder, C. L., Malys, N., Dunn, W. B., Murabito, E., Swainston, N., Dada, J. O., Khan, F., Pir, P., Simeonidis, E., Spasic, I., Wishart, J., Weichart, D., Hayes, N. W., Jameson, D., Broomhead, D. S., Oliver, S. G., Gaskell, S. J., McCarthy, J. E. G., Paton, N. W., Westerhoff, H. V., Kell, D. B., and Mendes, P. (2013) A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS Lett.* **587**, 2832–2841
54. Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P. Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., and Tomita, M. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* **316**, 593–597
55. Berthoumieux, S., Brilli, M., de Jong, H., Kahn, D., and Cinquemani, E. (2011) Identification of metabolic network models from incomplete high-throughput datasets. *Bioinformatics* **27**, i186–95